Sensitivity of uncertainty quantification and Bayesian inverse problems

Joint work with Oliver Ernst and Alois Pichler (TU Chemnitz)

June 8th, 2022
Outline

1. Uncertainty Propagation
2. Risk Functionals
3. Bayesian Inversion
- Consider general PDE with solution $u \in \mathcal{U}$ and coefficient(s) $a \in \mathcal{A}$ given by
  \[ \mathcal{F}(u, a) = 0 \]

- **Running example**: Elliptic diffusion equation on compact $D \subset \mathbb{R}^2$,
  \[ -\nabla \cdot (e^{a} \nabla u) = f, \quad u|_{\partial D} \equiv 0 \]
  with $\mathcal{U} = H^1_0(D)$ and $\mathcal{A} = L^\infty(D)$ (weak form)
Consider general PDE with solution \( u \in \mathcal{U} \) and coefficient(s) \( a \in \mathcal{A} \) given by
\[
\mathcal{F}(u, a) = 0
\]

**Running example**: Elliptic diffusion equation on compact \( D \subseteq \mathbb{R}^2 \),
\[
-\nabla \cdot (e^a \nabla u) = f, \quad u|_{\partial D} \equiv 0
\]
with \( \mathcal{U} = H^1_0(D) \) and \( \mathcal{A} = L^\infty(D) \) (weak form)

Often (log diffusion) coefficient \( a \) not precisely known in practice
Consider general PDE with solution \( u \in \mathcal{U} \) and coefficient(s) \( a \in \mathcal{A} \) given by
\[
\mathcal{F}(u, a) = 0
\]

**Running example:** Elliptic diffusion equation on compact \( D \subseteq \mathbb{R}^2 \),
\[
-\nabla \cdot (e^a \nabla u) = f, \quad u \big|_{\partial D} \equiv 0
\]
with \( \mathcal{U} = H^1_0(D) \) and \( \mathcal{A} = L^\infty(D) \) (weak form)

Often (log diffusion) coefficient \( a \) not precisely known in practice

**UQ approach:**
1. Describe uncertainty about \( a \) by probability measure \( \mu \) on \( \mathcal{A} \)
Consider general PDE with solution $u \in \mathcal{U}$ and coefficient(s) $a \in \mathcal{A}$ given by

$$\mathcal{F}(u, a) = 0$$

**Running example**: Elliptic diffusion equation on compact $D \subseteq \mathbb{R}^2$,

$$-\nabla \cdot (e^a \nabla u) = f, \quad u|_{\partial D} \equiv 0$$

with $\mathcal{U} = H^1_0(D)$ and $\mathcal{A} = L^\infty(D)$ (weak form)

Often (log diffusion) coefficient $a$ not precisely known in practice

**UQ approach**:

1. Describe uncertainty about $a$ by probability measure $\mu$ on $\mathcal{A}$
2. Compute resulting probability distribution $\nu$ on $\mathcal{U}$ of random solution $u$
Consider general PDE with solution $u \in \mathcal{U}$ and coefficient(s) $a \in \mathcal{A}$ given by

$$
\mathcal{F}(u, a) = 0
$$

**Running example**: Elliptic diffusion equation on compact $D \subseteq \mathbb{R}^2$,

$$
-\nabla \cdot (e^a \nabla u) = f, \quad u|_{\partial D} \equiv 0
$$

with $\mathcal{U} = H^1_0(D)$ and $\mathcal{A} = L^\infty(D)$ (weak form)

Often (log diffusion) coefficient $a$ not precisely known in practice

**UQ approach**:  

1. Describe uncertainty about $a$ by probability measure $\mu$ on $\mathcal{A}$  
2. Compute resulting probability distribution $\nu$ on $\mathcal{U}$ of random solution $u$ or quantity of interest $q(u)$ where $q : \mathcal{U} \to \mathbb{R}$
UQ Scheme

Distribution of $\mu$ often obtained by estimation or subjective knowledge

Question: Can we control the effect of perturbations of $\mu$ on output distribution $\nu$?

PDE

$F(u,a) = 0$
UQ Scheme

But: Distribution $\mu$ of $a$ often obtained by estimation or subjective knowledge.

PDE

$F(u,a) = 0$
But: Distribution $\mu$ of $a$ often obtained by estimation or subjective knowledge.
**But:** Distribution $\mu$ of $a$ often obtained by estimation or subjective knowledge

**Question**

Can we control the effect of perturbations of $\mu$ on output distribution $\nu$?
Simulate groundwater flow, here: at deep geological repository

Computational model given by PDE

\[-\nabla \cdot (e^a \nabla u) = f\]

Available data:
(a) log conductivity
(b) pressure head

Of interest: exit time of accidentally released radionuclides

UQ approach:
1. Model uncertain spatially varying log conductivity by Gaussian process
2. Use available data to estimate stochastic model for \(a\)
3. Compute resulting distribution of exit times \(q_{exit}\)

source: Sandia National Labs
Motivational Example

- Simulate groundwater flow, here: at deep geological repository
- Computational model given by PDE
  \[-\nabla \cdot (e^a \nabla u) = f\]
- Available data: \(\approx 60\) measurements of
  (a) log conductivity \(a\)
  (b) pressure head \(u\)
Motivational Example

- Simulate groundwater flow, here: at deep geological repository
- Computational model given by PDE
  \[-\nabla \cdot (e^a \nabla u) = f\]
- Available data: \(\approx 60\) measurements of
  (a) log conductivity \(a\)
  (b) pressure head \(u\)
- Of interest: exit time \(q_{\text{exit}}\) of accidentally released radionuclides
Simulate groundwater flow, here: at deep geological repository

Computational model given by PDE

\[-\nabla \cdot (e^a \nabla u) = f\]

Available data: \(\approx 60\) measurements of
(a) log conductivity \(a\)
(b) pressure head \(u\)

Of interest: exit time \(q_{\text{exit}}\) of accidentally released radionuclides

**UQ approach:**

1. Model uncertain spatially varying log conductivity \(a\) by Gaussian process
2. Use available data to estimate stochastic model for \(a\)
3. Compute resulting distribution of exit times \(q_{\text{exit}}\)
Gaussian random fields

- We assume $a \sim N(m, c)$ with mean and covariance function $m$ and $c$

$$
E[a(x)] = m(x),
$$
$$
\text{Cov}[a(x), a(y)] = c(x, y),
$$
$$
a(x) \sim N(m(x), c(x, x))
$$
Gaussian random fields

- We assume $a \sim N(m, c)$ with mean and covariance function $m$ and $c$

- Common parametrized class: Matérn covariance functions

$$c \sigma^2, \rho, k + \frac{1}{2} (x, y) := \sigma^2 \ e^{-\frac{\sqrt{2k+1}}{\rho} |x-y|} \ P_k \left( \frac{\sqrt{2k+1}}{\rho} |x-y| \right)$$

with variance $\sigma^2 > 0$, correlation length $\rho > 0$, smoothness $k + \frac{1}{2}$, $k \in \mathbb{N}_0$
Gaussian random fields

- We assume $a \sim \mathcal{N}(m, c)$ with mean and covariance function $m$ and $c$

- Common parametrized class: Matérn covariance functions

$$c_{\sigma^2, \rho, k+\frac{1}{2}}(x, y) := \sigma^2 \ e^{-\frac{\sqrt{2k+1}}{\rho} |x-y|} \ P_k \left( \frac{\sqrt{2k+1}}{\rho} |x - y| \right)$$

with variance $\sigma^2 > 0$, correlation length $\rho > 0$, smoothness $k + \frac{1}{2}$, $k \in \mathbb{N}_0$
Gaussian random fields

- We assume \( a \sim N(m, c) \) with mean and covariance function \( m \) and \( c \).

- Common parametrized class: Matérn covariance functions

\[
c_{\sigma^2, \rho, k + \frac{1}{2}}(x, y) := \sigma^2 e^{-\frac{\sqrt{2k+1}}{\rho} |x-y|} P_k \left( \frac{\sqrt{2k+1}}{\rho} |x-y| \right)
\]

with variance \( \sigma^2 > 0 \), correlation length \( \rho > 0 \), smoothness \( k + \frac{1}{2}, k \in \mathbb{N}_0 \).

- In practice, we obtain estimates \( \hat{\sigma}^2, \hat{\nu}, \hat{\rho} \) of the parameters given observational data \( a(x_j), j = 1, \ldots, n \) (e.g., by maximum-likelihood).

Motivational Question

How does estimation error or different choice of parameters, e.g., for \( \sigma^2 \), affect the output of the UQ analysis?
Solution Operator

• Consider solution operator $S: \mathcal{A} \to \mathcal{U}$ of PDE mapping coefficient $a$ to unique solution $u$ of $\mathcal{F}(u, a) = 0$
Consider solution operator \( S: \mathcal{A} \rightarrow \mathcal{U} \) of PDE mapping coefficient \( a \) to unique solution \( u \) of \( \mathcal{F}(u, a) = 0 \).

\[ \Rightarrow \text{Distribution } \nu \text{ of random solution } u \text{ is pushforward measure} \]

\[ \nu = S_*\mu, \quad \nu(A) = \mu(S^{-1}(A)), \quad A \subseteq \mathcal{U} \]
Consider solution operator \( S: \mathcal{A} \rightarrow \mathcal{U} \) of PDE mapping coefficient \( a \) to unique solution \( u \) of \( \mathcal{F}(u, a) = 0 \)

⇒ Distribution \( \nu \) of random solution \( u \) is pushforward measure

\[
\nu = S_*\mu, \quad \nu(A) = \mu(S^{-1}(A)), \quad A \subseteq \mathcal{U}
\]

Often \( S \) (locally) Lipschitz: with monotonically increasing \( \text{Lip}_S: \mathbb{R}_+ \rightarrow \mathbb{R}_+ \)

\[
\|S(a) - S(\tilde{a})\|_\mathcal{U} \leq \text{Lip}_S(r) \|a - \tilde{a}\|_\mathcal{A} \quad \forall \|a\|_\mathcal{A}, \|\tilde{a}\|_\mathcal{A} \leq r
\]
Solution Operator

- Consider solution operator $S: \mathcal{A} \to \mathcal{U}$ of PDE mapping coefficient $a$ to unique solution $u$ of $\mathcal{F}(u, a) = 0$

$\Rightarrow$ Distribution $\nu$ of random solution $u$ is pushforward measure

$$\nu = S_* \mu, \quad \nu(A) = \mu(S^{-1}(A)), \quad A \subseteq \mathcal{U}$$

- Often $S$ (locally) Lipschitz: with monotonically increasing $\text{Lip}_S: \mathbb{R}_+ \to \mathbb{R}_+$

$$\|S(a) - S(\widehat{a})\|_{\mathcal{U}} \leq \text{Lip}_S(r) \|a - \widehat{a}\|_{\mathcal{A}} \quad \forall \|a\|_{\mathcal{A}}, \|\widehat{a}\|_{\mathcal{A}} \leq r$$

- Running example: For elliptic problem $-\nabla \cdot (e^a \nabla u) = f$ we have

$$\|u - \widehat{u}\|_{H^1_0(D)} = \|S(a) - S(\widehat{a})\|_{H^1_0(D)} \leq c_f e^{3r} \|a - \widehat{a}\|_{L^\infty(D)}$$
Consider solution operator \( S : \mathcal{A} \rightarrow \mathcal{U} \) of PDE mapping coefficient \( a \) to unique solution \( u \) of \( \mathcal{F}(u, a) = 0 \)

\[ \Rightarrow \text{Distribution } \nu \text{ of random solution } u \text{ is pushforward measure} \]

\[ \nu = S_* \mu, \quad \nu(A) = \mu(S^{-1}(A)), \quad A \subseteq \mathcal{U} \]

Often \( S \) (locally) Lipschitz: with monotonically increasing \( \text{Lip}_S : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \)

\[ \|S(a) - S(\widehat{a})\|_\mathcal{U} \leq \text{Lip}_S(r) \|a - \widehat{a}\|_\mathcal{A} \quad \forall \|a\|_\mathcal{A}, \|\widehat{a}\|_\mathcal{A} \leq r \]

**Running example:** For elliptic problem \(-\nabla \cdot (e^{a} \nabla u) = f\) we have

\[ \|u - \widehat{u}\|_{H^1_0(D)} = \|S(a) - S(\widehat{a})\|_{H^1_0(D)} \leq c_f \ e^{3r} \|a - \widehat{a}\|_{L^\infty(D)} \]

Does Lipschitz continuity of \( S \) yield Lipschitz continuity of \( \mu \mapsto S_* \mu \)?
Sensitivity in TV Distance

- Consider total variation distance

\[ d_{TV} (\mu, \hat{\mu}) = \sup_{A \subseteq \mathcal{X}} |\mu(A) - \hat{\mu}(A)| \]
Sensitivity in TV Distance

- Consider total variation distance

\[ d_{TV} (\mu, \widehat{\mu}) = \sup_{A \subseteq \mathcal{A}} |\mu(A) - \widehat{\mu}(A)| \]

- Then, for any measurable \( S: \mathcal{A} \rightarrow \mathcal{U} \) we have global Lipschitz continuity of \( \mu \mapsto S_\ast \mu \):

\[ d_{TV} (S_\ast \mu, S_\ast \widehat{\mu}) \leq d_{TV} (\mu, \widehat{\mu}), \]

But:

TV distance not suited for measures on infinite-dimensional spaces, e.g., for Gaussian measures associated to Gaussian processes we have

\[ d_{TV} (\mathcal{N}(m, C), \mathcal{N}(m, \sigma^2 C)) = 1 \] if \( \sigma \neq 1 \) – i.e., any estimation error in variance parameter \( \sigma^2 \) yields maximal distance
Sensitivity in TV Distance

- Consider total variation distance

\[ d_{TV} (\mu, \widehat{\mu}) = \sup_{A \subseteq \mathcal{A}} |\mu(A) - \widehat{\mu}(A)| \]

- Then, for any measurable \( S: \mathcal{A} \rightarrow \mathcal{U} \) we have global Lipschitz continuity of \( \mu \mapsto S_\ast \mu \):

\[ d_{TV} (S_\ast \mu, S_\ast \widehat{\mu}) \leq d_{TV} (\mu, \widehat{\mu}) \]

- But: TV distance not suited for measures on infinite-dimensional spaces,
Sensitivity in TV Distance

- Consider total variation distance

\[ d_{\text{TV}}(\mu, \widehat{\mu}) = \sup_{A \subseteq \mathcal{A}} |\mu(A) - \widehat{\mu}(A)| \]

- Then, for any measurable \( S: \mathcal{A} \to \mathcal{U} \) we have global Lipschitz continuity of \( \mu \mapsto S_* \mu \):

\[ d_{\text{TV}}(S_* \mu, S_* \widehat{\mu}) \leq d_{\text{TV}}(\mu, \widehat{\mu}), \]

- But: TV distance not suited for measures on infinite-dimensional spaces, e.g., for Gaussian measures associated to Gaussian processes we have

\[ d_{\text{TV}}(N(m, C), N(m, \sigma^2 C)) = 1 \quad \text{if } \sigma \neq 1, \]

i.e., any estimation error in variance parameter \( \sigma^2 \) yields maximal distance
Instead, we consider the $p$-Wasserstein distance

$$W_p(\mu, \hat{\mu}) := \inf_{X \sim \mu, \hat{X} \sim \hat{\mu}} E \left[ \|X - \hat{X}\|^p \right]^{1/p}, \quad p \geq 1.$$
Wasserstein Distance

Instead, we consider the $p$-Wasserstein distance

$$ W_p(\mu, \hat{\mu}) := \inf_{X \sim \mu, \hat{X} \sim \hat{\mu}} \mathbb{E} \left[ \|X - \hat{X}\|^p \right]^{1/p}, \quad p \geq 1. $$

Also reasonable for measures which are singular w. r. t. each other

$$ W_p(\delta_a, \delta_{\hat{a}}) = \|a - \hat{a}\| $$
Wasserstein Distance

- Instead, we consider the $p$-Wasserstein distance

$$W_p(\mu, \hat{\mu}) := \inf_{X \sim \mu, \hat{X} \sim \hat{\mu}} \mathbb{E} \left[ \|X - \hat{X}\|^p \right]^{1/p}, \quad p \geq 1.$$ 

- Also reasonable for measures which are singular w. r. t. each other

$$W_p(\delta_a, \delta_{\hat{a}}) = \|a - \hat{a}\|$$

- Allows dual representation, e.g., for $W_1$ we have by Kantorovich–Rubinstein

$$W_1(\mu, \hat{\mu}) = \sup_{f : \mathcal{A} \to \mathbb{R}, \text{Lip}_f \leq 1} \left| \mathbb{E}_{\mu} [f] - \mathbb{E}_{\hat{\mu}} [f] \right|$$

\[ \text{Wasserstein Distance} \]
Wasserstein Distance

- Instead, we consider the $p$-Wasserstein distance

$$W_p(\mu, \hat{\mu}) := \inf_{X \sim \mu, \hat{X} \sim \hat{\mu}} \mathbb{E} \left[ \|X - \hat{X}\|^p \right]^{1/p}, \quad p \geq 1.$$  

- Also reasonable for measures which are singular w. r. t. each other

$$W_p(\delta_a, \delta_{\hat{a}}) = \|a - \hat{a}\|$$

- Allows dual representation, e.g., for $W_1$ we have by Kantorovich–Rubinstein

$$W_1(\mu, \hat{\mu}) = \sup_{f : \mathcal{A} \to \mathbb{R}, \text{Lip}_f \leq 1} \left| \mathbb{E}_{\mu} [f] - \mathbb{E}_{\hat{\mu}} [f] \right|$$

- $W_2$-distance of Gaussian measures explicitly known [Gelbrich, 1990],

$$W_2 \left( N(m, C), N(\hat{m}, \hat{C}) \right)^2 = \|m - \hat{m}\|^2 + \text{tr } C + \text{tr } \hat{C} - 2 \text{tr} \left( \sqrt{C} \hat{C} \sqrt{C} \right)^{1/2},$$

and, e.g., $W_p \left( N(m, \sigma^2 C), N(m, \hat{\sigma}^2 C) \right) \leq |\sigma - \hat{\sigma}|$
Sensitivity in Wasserstein Distance

- [Ernst, Pichler, S., 2020]: If $S: \mathcal{A} \to \mathcal{U}$ is globally Lipschitz with Lipschitz constant $\text{Lip}_S$, then for any probability measures $\mu, \tilde{\mu}$ on $\mathcal{A}$ we have

$$W_p (S_*\mu, S_*\tilde{\mu}) \leq \text{Lip}_S \ W_p (\mu, \tilde{\mu}), \quad p \geq 1.$$
Sensitivity in Wasserstein Distance

- [Ernst, Pichler, S., 2020]: If $S: \mathcal{A} \rightarrow \mathcal{U}$ is globally Lipschitz with Lipschitz constant $\text{Lip}_S$, then for any probability measures $\mu, \mu\hat{}$ on $\mathcal{A}$ we have

$$W_p (S_\ast \mu, S_\ast \mu\hat{}) \leq \text{Lip}_S \ W_p (\mu, \mu\hat{}), \quad p \geq 1.$$ 

- Applicable if $S$ is bounded and linear, e. g.,

$$f \mapsto u = S(f) \quad \text{for PDE} \quad -\nabla \cdot (e^a \nabla u) = f$$
Sensitivity in Wasserstein Distance

- [Ernst, Pichler, S., 2020]: If $S : \mathcal{A} \rightarrow \mathcal{U}$ is globally Lipschitz with Lipschitz constant $\text{Lip}_S$, then for any probability measures $\mu, \widehat{\mu}$ on $\mathcal{A}$ we have

$$W_p (S_*\mu, S_*\widehat{\mu}) \leq \text{Lip}_S \ W_p (\mu, \widehat{\mu}), \quad p \geq 1.$$ 

- Applicable if $S$ is bounded and linear, e. g.,

$$f \mapsto u = S(f) \quad \text{for PDE} \quad - \nabla \cdot (e^a \nabla u) = f$$

- **But:** For general locally Lipschitz forward maps $S$ we do not obtain global Lipschitz continuity of $\mu \mapsto S_*\mu$
Sensitivity in Wasserstein Distance

- [Ernst, Pichler, S., 2020]: If $S: \mathcal{A} \rightarrow \mathcal{U}$ is globally Lipschitz with Lipschitz constant $\text{Lip}_S$, then for any probability measures $\mu, \tilde{\mu}$ on $\mathcal{A}$ we have

$$W_p \left( S_*\mu, S_*\tilde{\mu} \right) \leq \text{Lip}_S \, W_p \left( \mu, \tilde{\mu} \right), \quad p \geq 1.$$ 

- Applicable if $S$ is bounded and linear, e.g.,

$$f \mapsto u = S(f) \quad \text{for PDE} \quad -\nabla \cdot (e^a \nabla u) = f$$

- **But**: For general locally Lipschitz forward maps $S$ we do not obtain global Lipschitz continuity of $\mu \mapsto S_*\mu$

**Example**: $\mu = N(0, 1)$, $\mu_\epsilon = N(0, 1 + \epsilon)$ and $S(x) := e^x$, $x \in \mathbb{R}$, then

$$\lim_{\epsilon \to +\infty} \frac{W_p \left( S_*\mu, S_*\mu_\epsilon \right)}{W_p \left( \mu, \mu_\epsilon \right)} = +\infty$$
However, we can recover local Lipschitz continuity of $\mu \mapsto S_* \mu$ under suitable assumptions:
However, we can recover local Lipschitz continuity of $\mu \mapsto S*\mu$ under suitable assumptions:

**Theorem ([Ernst, Pichler, S., 2020])**

Let $S : \mathcal{A} \rightarrow \mathcal{U}$ be locally Lipschitz,

$$
\|S(a) - S(\hat{a})\|_{\mathcal{U}} \leq \text{Lip}_S(r) \|a - \hat{a}\|_{\mathcal{A}} \quad \forall \|a\|_{\mathcal{A}}, \|\hat{a}\|_{\mathcal{A}} \leq r.
$$
However, we can recover local Lipschitz continuity of $\mu \mapsto S_\ast \mu$ under suitable assumptions:

**Theorem ([Ernst, Pichler, S., 2020])**

Let $S: \mathcal{A} \to \mathcal{U}$ be locally Lipschitz,

$$
\|S(a) - S(\hat{a})\|_\mathcal{U} \leq \text{Lip}_S(r) \|a - \hat{a}\|_\mathcal{A} \quad \forall \|a\|_\mathcal{A}, \|\hat{a}\|_\mathcal{A} \leq r.
$$

Then for any $\mu, \tilde{\mu}$ with

$$
E_\mu \left[ \text{Lip}_S^{2p}(\|a\|_\mathcal{A}) \right], \quad E_{\tilde{\mu}} \left[ \text{Lip}_S^{2p}(\|a\|_\mathcal{A}) \right] \leq C < \infty
$$
However, we can recover local Lipschitz continuity of $\mu \mapsto S_*\mu$ under suitable assumptions:

**Theorem ([Ernst, Pichler, S., 2020])**

Let $S: \mathcal{A} \rightarrow \mathcal{U}$ be locally Lipschitz,

\[ \|S(a) - S(\hat{a})\|_{\mathcal{U}} \leq \text{Lip}_S(r) \|a - \hat{a}\|_{\mathcal{A}} \quad \forall \|a\|_{\mathcal{A}}, \|\hat{a}\|_{\mathcal{A}} \leq r. \]

Then for any $\mu, \tilde{\mu}$ with

\[ E_\mu \left[ \text{Lip}^2_S (\|a\|_{\mathcal{A}}) \right], \quad E_{\tilde{\mu}} \left[ \text{Lip}^2_S (\|a\|_{\mathcal{A}}) \right] \leq C < \infty \]

we have

\[ W_p (S_*\mu, S_*\tilde{\mu}) \leq 2C^{1/2p} W_{2p} (\mu, \tilde{\mu}). \]
However, we can recover local Lipschitz continuity of $\mu \mapsto S_\mu$ under suitable assumptions:

**Theorem ([Ernst, Pichler, S., 2020])**

Let $S: \mathcal{A} \to \mathcal{U}$ be locally Lipschitz,

$$\|S(a) - S(\hat{a})\|_\mathcal{U} \leq \text{Lip}_S(r) \|a - \hat{a}\|_\mathcal{A} \quad \forall \|a\|_\mathcal{A}, \|\hat{a}\|_\mathcal{A} \leq r.$$ 

Then for any $\mu, \hat{\mu}$ with

$$E_\mu \left[ \text{Lip}_S^{2p}(\|a\|_\mathcal{A}) \right], \ E_{\hat{\mu}} \left[ \text{Lip}_S^{2p}(\|a\|_\mathcal{A}) \right] \leq C < \infty$$

we have

$$W_p(S_\mu, S_{\hat{\mu}}) \leq 2C^{1/2p} W_{2p}(\mu, \hat{\mu}).$$

Which measures $\mu, \hat{\mu}$ satisfy the integrability assumption for $\text{Lip}_S(r) \in \Theta(e^{\beta r})$?
Recall Gaussian random fields $a \sim \mathcal{N}(m, c)$ with continuous mean function $m \in C(D)$, $D \subset \mathbb{R}^d$ compact, and Matérn covariance functions

$$c_{\sigma^2, \rho, k+\frac{1}{2}}(x, y) := \sigma^2 e^{-\frac{\sqrt{2k+1}}{\rho} |x-y|} \frac{k!}{(2k)!} \sum_{i=0}^{k} \frac{(k+i)!}{i!(k-i)!} \left(2 \frac{\sqrt{2k+1}}{\rho} |x-y| \right)^{k-i}$$
Special case: Gaussian random fields

- Recall Gaussian random fields $a \sim \text{N}(m, c)$ with continuous mean function $m \in C(D)$, $D \subset \mathbb{R}^d$ compact, and Matérn covariance functions

$$c_{\sigma^2, \rho, k + \frac{1}{2}}(x, y) := \sigma^2 e^{-\frac{\sqrt{2k+1}}{\rho} |x-y|} \frac{k!}{(2k)!} \sum_{i=0}^{k} \frac{(k+i)!}{i!(k-i)!} \left(2 \frac{\sqrt{2k+1}}{\rho} |x-y|\right)^{k-i}$$

- We consider the following subclass $\mathcal{G}$ of Gaussian measures on $C(D)$

$$\mathcal{G} = \mathcal{G}(\mathcal{M}, \mathcal{C}) := \{\text{N}(m, c) : m \in \mathcal{M}, c \in \mathcal{C}\}$$

$$\mathcal{M} = \{m : \|m\|_{C(D)} \leq r M\}$$

$$\mathcal{C} = \left\{ c_{\sigma^2, \rho, k + \frac{1}{2}} : \sigma \leq \sigma_{\text{max}}, \ \rho \geq \rho_{\text{min}}, k \in \{0, \ldots, k_{\text{max}}\} \right\}$$
[Ernst, Pichler, S., 2020]: By means of Fernique's theorem and Dudley's entropy bound we have

\[
\sup_{\mu \in \mathcal{G}} E_\mu \left[ \exp \left( \beta \|a\|_{C(D)} \right) \right] < \infty, \quad \beta > 0.
\]
By means of Fernique’s theorem and Dudley’s entropy bound we have

\[ \sup_{\mu \in \mathcal{G}} \mathbb{E}_\mu \left[ \exp \left( \beta \|a\|_{C(D)} \right) \right] < \infty, \quad \beta > 0. \]

**Theorem ([Ernst, Pichler, S., 2020])**

Consider \( \mathcal{G} = \mathcal{G}(\mathcal{M}, C) \) and locally Lipschitz \( S : C(D) \rightarrow \mathcal{U} \) with \( \text{Lip}_S(r) \in O(e^{\beta r}) \) for a \( \beta > 0 \).
• [Ernst, Pichler, S., 2020]: By means of Fernique’s theorem and Dudley’s entropy bound we have

$$\sup_{\mu \in \mathcal{G}} \mathbb{E}_\mu \left[ \exp \left( \beta \|a\|_{C(D)} \right) \right] < \infty, \quad \beta > 0.$$ 

**Theorem ([Ernst, Pichler, S., 2020])**

Consider $\mathcal{G} = \mathcal{G}(\mathcal{M}, C)$ and locally Lipschitz $S : C(D) \to \mathcal{U}$ with $\text{Lip}_S(r) \in \mathcal{O}(e^{\beta r})$ for a $\beta > 0$. Then, there exists a constant $C = C(\mathcal{G}) < \infty$ such that

$$W_p (S_*\mu, S_*\hat{\mu}) \leq C \ W_{2p} (\mu, \hat{\mu}) \quad \forall \mu, \hat{\mu} \in \mathcal{G}.$$
• [Ernst, Pichler, S., 2020]: By means of Fernique’s theorem and Dudley’s entropy bound we have

\[ \sup_{\mu \in \mathcal{G}} E_\mu \left[ \exp \left( \beta \|a\|_{C(D)} \right) \right] < \infty, \quad \beta > 0. \]

**Theorem ([Ernst, Pichler, S., 2020])**

Consider \( \mathcal{G} = \mathcal{G}(M, C) \) and locally Lipschitz \( S : C(D) \to \mathcal{U} \) with \( \text{Lip}_S(r) \in \Theta(e^{\beta r}) \) for a \( \beta > 0 \). Then, there exists a constant \( C = C(\mathcal{G}) < \infty \) such that

\[ W_p(S_*\mu, S_*\hat{\mu}) \leq C \ W_{2p}(\mu, \hat{\mu}) \quad \forall \mu, \hat{\mu} \in \mathcal{G}. \]

• **Example:** For elliptic problem \(-\nabla \cdot (e^a \nabla u) = f\) with lognormal diffusion coefficients we have for \( \mu = N(m, c \sigma^2, \rho, k + \frac{1}{2}) \), \( \hat{\mu} = N(m, \tilde{c} \tilde{\sigma}^2, \rho, k + \frac{1}{2}) \)

\[ W_p(S_*\mu, S_*\hat{\mu}) \leq C_{\sigma_{\text{max}}} |\sigma - \tilde{\sigma}| \quad \forall \sigma, \tilde{\sigma} \leq \sigma_{\text{max}} \]
Risk Functionals

- Consider now scalar quantity of interest $q : \mathcal{U} \to \mathbb{R}$ of solution $u$ of PDE.
Consider now scalar quantity of interest \( q : \mathcal{U} \rightarrow \mathbb{R} \) of solution \( u \) of PDE

- Tool to evaluate uncertainty about quantity \( q(u) \): risk functionals
Risk Functionals

- Consider now scalar \textit{quantity of interest} $q: \mathcal{U} \rightarrow \mathbb{R}$ of solution $u$ of PDE

- Tool to evaluate uncertainty about quantity $q(u)$: risk functionals

- Risk functionals $R$ assign real numbers $R(X) \in \mathbb{R}$ to (real-valued) random variables $X$ which quantify the risk associated with their random outcomes
Risk Functionals

- Consider now scalar quantity of interest \( q : \mathcal{U} \to \mathbb{R} \) of solution \( u \) of PDE

- Tool to evaluate uncertainty about quantity \( q(u) \): risk functionals

- Risk functionals \( R \) assign real numbers \( R(X) \in \mathbb{R} \) to (real-valued) random variables \( X \) which quantify the risk associated with their random outcomes

Examples:
- Expectation: \( R(X) = \mathbb{E}[X] \)
Consider now scalar \( q : \mathcal{U} \rightarrow \mathbb{R} \) of solution \( u \) of PDE

Tool to evaluate uncertainty about quantity \( q(u) \): risk functionals

Risk functionals \( R \) assign real numbers \( R(X) \in \mathbb{R} \) to (real-valued) random variables \( X \) which quantify the risk associated with their random outcomes

Examples:

- Expectation: \( R(X) = E[X] \)

- Value-at-Risk (VaR): \( R(X) := F_X^{-1}(1 - \alpha), \quad \alpha \in (0, 1) \)
Consider now scalar quantity of interest $q: \mathcal{U} \to \mathbb{R}$ of solution $u$ of PDE

Tool to evaluate uncertainty about quantity $q(u)$: risk functionals

Risk functionals $R$ assign real numbers $R(X) \in \mathbb{R}$ to (real-valued) random variables $X$ which quantify the risk associated with their random outcomes

Examples:

- Expectation: $R(X) = E[X]$

- Value-at-Risk (VaR): $R(X) := F_X^{-1}(1 - \alpha)$, $\alpha \in (0, 1)$

- Average Value-at-Risk (AVaR): $R(X) = \frac{1}{\alpha} \int_{1-\alpha}^{1} F_X^{-1}(t)dt$, $\alpha \in (0, 1)$
Consider now scalar quantity of interest \( q : \mathcal{U} \rightarrow \mathbb{R} \) of solution \( u \) of PDE.

Tool to evaluate uncertainty about quantity \( q(u) \): risk functionals

Risk functionals \( R \) assign real numbers \( R(X) \in \mathbb{R} \) to (real-valued) random variables \( X \) which quantify the risk associated with their random outcomes.

Examples:

- Expectation: \( R(X) = \mathbb{E}[X] \)

- Value-at-Risk (VaR): \( R(X) := F_X^{-1}(1 - \alpha), \quad \alpha \in (0, 1) \)

- Average Value-at-Risk (AVaR): \( R(X) = \frac{1}{\alpha} \int_{1-\alpha}^{1} F_X^{-1}(t) \, dt, \quad \alpha \in (0, 1) \)

- Spectral risk functional: \( R(X) = \int_0^1 w(t) \, F_X^{-1}(t) \, dt, \quad w \in L^1(\mathbb{R}_+) \)
Risk Functionals – Illustration

Goal

Control effect of underlying distribution $u$ on risk value $R$.
Risk Functionals – Illustration

Goal
Control effect of underlying distribution $u \sim v$ on risk value $R(q(u))$
Coherent Risk Functionals [Artzner et al., 1997]

- Common class of risk functionals which are
  1. **monotone:** \( X \leq Y \implies R(X) \leq R(Y) \)
  2. **cash-invariant:** \( R(X - c) = R(X) - c \) for any \( c \in \mathbb{R} \)
  3. **subadditive:** \( R(X + Y) \leq R(X) + R(Y) \)
  4. **positive homogeneous:** \( R(\lambda X) = \lambda R(X) \) for any \( \lambda > 0 \)

Spectral risk functionals such as AVaR are coherent, but VaR is not.
Coherent Risk Functionals  [Artzner et al., 1997]

- Common class of risk functionals which are

  1. **monotone:** \( X \preceq a.s. Y \implies R(X) \leq R(Y) \)

  2. **cash-invariant:** \( R(X - c) = R(X) - c \) for any \( c \in \mathbb{R} \)

  3. **subadditive:** \( R(X + Y) \leq R(X) + R(Y) \)

  4. **positive homogeneous:** \( R(\lambda X) = \lambda R(X) \) for any \( \lambda > 0 \)

- Spectral risk functionals such as AVaR are coherent, but VaR is not
Coherent Risk Functionals [Artzner et al., 1997]

- Common class of risk functionals which are
  1. **monotone:**  $X \lesssim Y \Rightarrow R(X) \leq R(Y)$
  2. **cash-invariant:**  $R(X - c) = R(X) - c$ for any $c \in \mathbb{R}$
  3. **subadditive:**  $R(X + Y) \leq R(X) + R(Y)$
  4. **positive homogeneous:**  $R(\lambda X) = \lambda R(X)$ for any $\lambda > 0$

- Spectral risk functionals such as AVaR are coherent, but VaR is not

**Dual representation**

By means of the Fenchel–Moreau theorem

$$R(X) = \sup_{H \in \mathcal{H}} E[H X], \quad \mathcal{H} \subseteq \{H : H \geq 0 \text{ a. s. and } E[H] = 1\},$$
i.e., $H$ basically represent probability density functions.
Coherent Risk Functionals [Artzner et al., 1997]

- Common class of risk functionals which are
  1. **monotone:** \( X \leq a.s. Y \implies R(X) \leq R(Y) \)
  2. **cash-invariant:** \( R(X - c) = R(X) - c \) for any \( c \in \mathbb{R} \)
  3. **subadditive:** \( R(X + Y) \leq R(X) + R(Y) \)
  4. **positive homogeneous:** \( R(\lambda X) = \lambda R(X) \) for any \( \lambda > 0 \)

- Spectral risk functionals such as AVaR are coherent, but VaR is not

---

**Theorem ([Ernst, Pichler, S., 2020])**

For Hölder-continuous quantity \( q : \mathcal{U} \to \mathbb{R} \), i.e., \( |q(u) - q(\widehat{u})| \leq C_q \| u - \widehat{u} \|_{\mathcal{U}}^\beta \), \( \beta > 0 \),
Coherent Risk Functionals [Artzner et al., 1997]

- Common class of risk functionals which are
  1. **monotone:** \( X \leq Y \Rightarrow R(X) \leq R(Y) \)
  2. **cash-invariant:** \( R(X - c) = R(X) - c \) for any \( c \in \mathbb{R} \)
  3. **subadditive:** \( R(X + Y) \leq R(X) + R(Y) \)
  4. **positive homogeneous:** \( R(\lambda X) = \lambda R(X) \) for any \( \lambda > 0 \)

- Spectral risk functionals such as AVaR are coherent, but VaR is not

**Theorem ([Ernst, Pichler, S., 2020])**

For Hölder-continuous quantity \( q : \mathcal{U} \to \mathbb{R} \), i.e., \(|q(u) - q(\hat{u})| \leq C_q \|u - \hat{u}\|_{\mathcal{U}}^\beta\), \( \beta > 0 \), we have for any coherent risk functional \( R \) that

\[
|R(q(u)) - R(q(\hat{u}))| \leq C_{R,p,q} \ W_p (\nu, \hat{\nu})^\beta, \quad p \geq 1
\]

where \( u \sim \nu \) and \( \hat{u} \sim \hat{\nu} \).
Can combine now previous results and obtain

\[ \text{Corollary (Ernst, Pichler, S., 2020)} \]

For Hölder-continuous \( q \): \( U \to R \) and locally Lipschitz \( S \): \( A \to U \) we have for any spectral risk measures \( R \) and suitable measures \( \mu - \beta \mu \) on \( A \)

\[ j R^{\frac{1}{2}} q j u \approx j C^{\frac{1}{2}} \beta \approx 1 - j \]

where \( u = S^{\frac{1}{2}} a \approx b \), \( a \approx \beta \) and \( b \approx \beta \).

Example: For elliptic problem \( r^{\frac{1}{2}} e a r u \approx f \) with lognormal diffusion coefficients we have for \( a \approx N^{\frac{1}{2}} m - c \sigma^2 - \rho - k \approx \frac{1}{2} \)]

\[ AVaR^{\frac{1}{2}} q AVaR^{\frac{1}{2}} b \approx j C^{\frac{1}{2}} \sigma_{\text{max}} j \sigma \approx \beta 8 \sigma - b \sigma \approx \sigma_{\text{max}} \]

for Hölder-continuous \( q \): \( H_{10}^{1/2} R \).
Can combine now previous results and obtain

**Corollary ([Ernst, Pichler, S., 2020])**

For Hölder-continuous $q : \mathcal{U} \to \mathbb{R}$ and locally Lipschitz $S : \mathcal{A} \to \mathcal{U}$ we have for any spectral risk measures $R$ and suitable measures $\mu, \hat{\mu}$ on $\mathcal{A}$

$$|R(q(u)) - R(q(\hat{u}))| \leq C_{q,w,p} \ W_{2p} (\mu, \hat{\mu})^{\beta}, \quad p \geq 1,$$

where $u = S(a)$, $a \sim \mu$, and $\hat{u} = S(\hat{a})$, $\hat{a} \sim \hat{\mu}$.
Sensitivity of Risk Functionals for random PDE

Can combine now previous results and obtain

**Corollary ([Ernst, Pichler, S., 2020])**

For Hölder-continuous \( q : \mathcal{U} \to \mathbb{R} \) and locally Lipschitz \( S : \mathcal{A} \to \mathcal{U} \) we have for any spectral risk measures \( R \) and suitable measures \( \mu, \mu \hat{\mu} \) on \( \mathcal{A} \)

\[
|R(q(u)) - R(q(\hat{u}))| \leq C_{q,w,p} W_{2p} (\mu, \mu \hat{\mu})^\beta, \quad p \geq 1,
\]

where \( u = S(a) \), \( a \sim \mu \), and \( \hat{u} = S(\hat{a}) \), \( \hat{a} \sim \hat{\mu} \).

**Example:** For elliptic problem \(-\nabla \cdot (e^a \nabla u) = f\) with lognormal diffusion coefficients we have for \( a \sim \mathcal{N}(m, c \sigma^2, \rho, k + \frac{1}{2}) \), \( \hat{a} \sim \mathcal{N}(m, c \hat{\sigma}^2, \rho, k + \frac{1}{2}) \)

\[
|\text{AVaR}(q(u)) - \text{AVaR}(q(\hat{u}))| \leq C_{\sigma_{\text{max}}} |\sigma - \hat{\sigma}|^\beta, \quad \forall \sigma, \hat{\sigma} \leq \sigma_{\text{max}}
\]

for Hölder-continuous \( q : H^1_0(D) \to \mathbb{R} \)
Bayesian Inverse Problems (BIP)

- UQ approach to inverse problem

\[ y = G(a) + \varepsilon, \quad G: \mathcal{A} \rightarrow \mathbb{R}^k, \quad \varepsilon \sim \mathcal{N}(0, \Sigma), \]

\[ \text{e.g., } G = O \circ S \text{ with observational map } O: \mathcal{U} \rightarrow \mathbb{R}^k \text{ applied to } u = S(a) \]
Bayesian Inverse Problems (BIP)

- **UQ approach to inverse problem**

\[ y = G(a) + \varepsilon, \quad G: \mathcal{A} \rightarrow \mathbb{R}^k, \quad \varepsilon \sim \mathcal{N}(0, \Sigma), \]

E.g., \( G = O \circ S \) with observational map \( O: \mathcal{Y} \rightarrow \mathbb{R}^k \) applied to \( u = S(a) \)

- **Update prior measure** \( \mu \) on \( \mathcal{A} \) for uncertain \( a \) by conditioning \( \mu \) on data
Bayesian Inverse Problems (BIP)

- **UQ approach to inverse problem**
  \[ y = G(a) + \epsilon, \quad G: \mathcal{A} \rightarrow \mathbb{R}^k, \quad \epsilon \sim \mathcal{N}(0, \Sigma), \]
e.g., \( G = O \circ S \) with observational map \( O: \mathcal{U} \rightarrow \mathbb{R}^k \) applied to \( u = S(a) \)

- **Update prior measure** \( \mu \) on \( \mathcal{A} \) for uncertain \( a \) by conditioning \( \mu \) on data

- **Bayes’ rule:** Posterior measure of \( a \sim \mu \) given data \( y = G(a) + \epsilon \) is

\[
\mu_\Phi(da) \propto e^{-\Phi(a)} \mu(da), \quad \Phi(a) := \frac{1}{2} \| y - G(a) \|_\Sigma^{-1}^2.
\]
Bayesian Inverse Problems (BIP)

- UQ approach to inverse problem

\[ y = G(a) + \varepsilon, \quad G: \mathcal{A} \to \mathbb{R}^k, \quad \varepsilon \sim \mathcal{N}(0, \Sigma), \]

\text{e.g., } G = O \circ S \text{ with observational map } O: \mathcal{U} \to \mathbb{R}^k \text{ applied to } u = S(a)

- Update prior measure \( \mu \) on \( \mathcal{A} \) for uncertain \( a \) by conditioning \( \mu \) on data

- Bayes’ rule: Posterior measure of \( a \sim \mu \) given data \( y = G(a) + \varepsilon \) is

\[
\mu_\Phi(da) \propto e^{-\Phi(a)} \mu(da), \quad \Phi(a) := \frac{1}{2} \| y - G(a) \|^2_{\Sigma^{-1}}.
\]

- BIP well-posed, i.e., local Lipschitz dependence of \( \mu_\Phi \) on data \( y \in \mathbb{R}^k \)
  [Stuart, 2010], [Hosseini, 2017], [Sullivan, 2017], [Latz, 2020],...
Bayesian Inverse Problems (BIP)

- **UQ approach to inverse problem**

  \[ y = G(a) + \varepsilon, \quad G: \mathcal{A} \to \mathbb{R}^k, \quad \varepsilon \sim \mathcal{N}(0, \Sigma), \]

  e.g., \( G = O \circ S \) with observational map \( O: \mathcal{U} \to \mathbb{R}^k \) applied to \( u = S(a) \)

- **Update prior measure** \( \mu \) on \( \mathcal{A} \) for uncertain \( a \) by conditioning \( \mu \) on data

- **Bayes’ rule**: Posterior measure of \( a \sim \mu \) given data \( y = G(a) + \varepsilon \) is

  \[
  \mu_\Phi(da) \propto e^{-\Phi(a)} \mu(da), \quad \Phi(a) := \frac{1}{2} \| y - G(a) \|_\Sigma^{-1}^2.
  \]

- **BIP well-posed, i.e., local Lipschitz dependence of** \( \mu_\Phi \) **on data** \( y \in \mathbb{R}^k \)
  
  [Stuart, 2010], [Hosseini, 2017], [Sullivan, 2017], [Latz, 2020], ...

- **Question**: How sensitively depends \( \mu_\Phi \) on (subjective) choice of \( \mu \)?
Theorem (informal, [S., 2020])

For $d$ being TV, Hellinger, or $1$-Wasserstein distance or KL divergence we have under suitable assumptions a locally Lipschitz continuity:

$$d^1_{\mu} \Phi - b \mu \leq C \Phi^1_r \leq d^1_{\mu} - b \mu - \gamma$$

But:

$$C \Phi^1_r \not\to 1$$

as data $y$ more informative, e.g., noise $\epsilon \to 0$.
Theorem (informal, [S., 2020])

For $d$ being TV, Hellinger, or 1-Wasserstein distance or KL divergence we have under suitable assumptions a locally Lipschitz continuity:

$$d(\mu_\Phi, \widehat{\mu}_\Phi) \leq C_\Phi(r) \ d(\mu, \widehat{\mu}), \quad \text{if} \ d(\mu, \widehat{\mu}) \leq r$$
Sensitivity of Bayesian Inversion

**Theorem (informal, [S., 2020])**

For \( d \) being TV, Hellinger, or 1-Wasserstein distance or KL divergence we have under suitable assumptions a locally Lipschitz continuity:

\[
d(\mu_\Phi, \hat{\mu}_\Phi) \leq C_\Phi(r) \ d(\mu, \hat{\mu}), \quad \text{if} \quad d(\mu, \hat{\mu}) \leq r
\]

**But:** \( C_\Phi(r) \to \infty \) as data \( y \) more informative, e.g., noise \( \varepsilon \to 0 \)
Theorem

1. If $\Phi: \mathcal{A} \to \mathbb{R}_+$ is continuous, we have continuity in $p$-Wasserstein distance, i.e.,

$$
\lim_{n \to \infty} W_p \left( \mu, \hat{\mu}^{(n)} \right) = 0 \quad \Rightarrow \quad \lim_{n \to \infty} W_p \left( \mu_{\Phi}, \hat{\mu}_{\Phi}^{(n)} \right) = 0, \quad p \geq 1.
$$
Theorem

1. If $\Phi: \mathcal{A} \rightarrow \mathbb{R}_+$ is continuous, we have continuity in $p$-Wasserstein distance, i.e.,

$$\lim_{n \rightarrow \infty} W_p \left( \mu, \hat{\mu}^{(n)} \right) = 0 \quad \Rightarrow \quad \lim_{n \rightarrow \infty} W_p \left( \mu_\Phi, \hat{\mu}_\Phi^{(n)} \right) = 0, \quad p \geq 1.$$

2. If $\mathcal{A}$ is bounded and $e^{-\Phi}: \mathcal{A} \rightarrow \mathbb{R}_+$ globally Lipschitz, then

$$W_1(\mu_\Phi, \hat{\mu}_\Phi) \leq \frac{C_\Phi}{Z^2} W_1(\mu, \hat{\mu}) \quad \forall \hat{\mu}: W_1(\mu, \hat{\mu}) \leq \frac{Z}{2 \text{Lip}_\Phi}$$

where $Z := \int e^{-\Phi} d\mu$ denotes normalizing constant for $\mu_\Phi$. 

Remark: [Diaconis & Freedman, 1986] studied Fréchet derivative $\partial T \Phi^{\cdot (n)}$ of mapping $T \Phi^{\cdot (n)} = \mu_\Phi$ w.r.t. TV distance topology and obtained $k \partial T \Phi^{\cdot (n)} \leq 1 \frac{Z}{\text{Lip}_\Phi}$. 


Wasserstein Distance

**Theorem**

1. If $\Phi : \mathcal{A} \rightarrow \mathbb{R}_+$ is continuous, we have continuity in $p$-Wasserstein distance, i.e.,

   \[
   \lim_{n \to \infty} W_p \left( \mu, \hat{\mu}^{(n)} \right) = 0 \quad \Rightarrow \quad \lim_{n \to \infty} W_p \left( \mu_{\Phi}, \hat{\mu}_{\Phi}^{(n)} \right) = 0, \quad p \geq 1.
   \]

2. If $\mathcal{A}$ is bounded and $e^{-\Phi} : \mathcal{A} \rightarrow \mathbb{R}_+$ globally Lipschitz, then

   \[
   W_1(\mu_{\Phi}, \hat{\mu}_{\Phi}) \leq \frac{C_{\Phi}}{Z^2} W_1(\mu, \hat{\mu}) \quad \forall \hat{\mu} : W_1(\mu, \hat{\mu}) \leq \frac{Z}{2 \text{Lip}_{\Phi}}
   \]

   where $Z := \int e^{-\Phi} d\mu$ denotes normalizing constant for $\mu_{\Phi}$.

**Remark:** [Diaconis & Freedman, 1986] studied Fréchet derivative $\partial T_{\Phi}(\mu)$ of mapping $T_{\Phi}(\mu) := \mu_{\Phi}$ w.r.t. TV distance topology and obtained

\[
\|\partial T_{\Phi}(\mu)\| \approx \frac{1}{Z}
\]
• Locally Lipschitz sensitivity of uncertainty propagation w.r.t. Wasserstein distance for locally Lipschitz forward maps

• Also locally Hölder sensitivity of risk functionals for Hölder-continuous quantities of interest

• Similar results for sensitivity of Bayesian inversion w.r.t. choice of prior (or perturbations of log-Likelihood $\Phi$)

More information:
